Close-down Report on (N16653) Software Effectiveness Metrics

September 25, 2015

Kenneth J. Roche[1]

The main goal of this project was to annually improve the computational science capabilities for a selected set of Scientific Discovery through Advanced Computing (SciDAC) and base applications from the U. S. Department of Energy (DOE) -in particular the Office of Science (SC). The focus was on the software implementation of the targeted application problems. A scientific baseline was stated per application and baseline problems were computed on targeted SC hardware platform(s) to indicate the existing capabilities of the software implementation. The process of enhancing the software applications focused on changes in data structures and algorithms, libraries, language and programming model enhancements, and compiler utilization with the goals of improved efficiency, scalability, and new results. Enhancements of the application's science and engineering capabilities focused on improved physical models, the problem representation, validity of inputs, and correctness of computed results. The enhanced applications were re-tested on the same target platform. A comparative analysis was made of the science and performance observables measured for the original and enhanced versions of the software to demonstrate and explain the enhancements. To conduct this work there was plenty of oversight from DOE headquarters because the results of the quarterly updates and final results were aligned with a DOE metric used by the US Office of Management and Budget executed by the Executive Office of the President of the US during the federal government's

[1]High Performance Computing Group, Pacific Northwest National Laboratory, 1 Battelle Road, Richland, Washington U.S.A. Email:`kenneth.roche@pnnl.gov`

budget request to the US Congress. The applications and targeted hardware platforms were selected by the Advanced Scientific Computing Advisory Committee (ASCAC) based on input from leadership at DOE headquarters and within the US national laboratory system. Full access to the Leadership Class supercomputers deployed and managed by the DOE SC Advanced Scientific Computing Research (ASCR) program was granted in order to do the stated work -an allocation and highest priority when needed was dedicated to this work on the targeted hardware platforms for the duration of the exercises. The project generated information quarterly for each application investigated about measured performance coupled to scientific achievement - software effectiveness. These results were used to satisfy the language of ASCR's annual performance goals and were led technically by the application PIs and Roche, and communicated to headquarters quarterly by Roche. These goals and the federal use of this work at the time are described in the appendix of this note. The language of the following software effectiveness metric was stated by then program manager at DOE SC ASCR, Dr. Daniel Hitchcock.

SC GG 3.1/2.5.2 *Improve computational science capabilities, defined as the average annual percentage increase in the computational effectiveness (either by simulating the same problem in less time or simulating a larger problem in the same time) of a subset of application codes. FY performance metric: efficiency measure is 100%.*

In practice, there was no interpretation for how to measure such a goal. This project introduced a structure for measuring and improving effective use of DOE's computing investments by determining the complexity affiliated with computing significant problems with specific algorithm and machine instances by direct measurement. Please refer to the final reports that were submitted to DOE each year for more details. The reports are listed here along with some other information excluding FY12 results. A couple of interesting things to note. This work and infrastructure remained effective in transitioning from the GigaFLOP to PetaFLOP investments in computing machines by DOE, was conducted by the same PI for three different ASCR associate directors (Dr. Edward Oliver, Dr. Michael Strayer, Dr. Daniel Hitchcock), was in response to two different base programs (MICS -Mathematical, Information, and Computational Sciences, and ASCR -Advanced Scientific Computing Research). The first open-science exascale machine has not been deployed at the time this project stopped.

climate research	4
condensed matter	4
fusion	5
high energy physics	3
nuclear	2
subsurface modeling	2
astrophysics	2
combustion chemistry	4
bioinformatics	1
math, data analytics	2
molecular dynamics, electronic structure	3
nuclear energy	1
Total	33

Cray	X1
	X1E
	XT3
	XT4
4-core	XT5
6-core	XT5
IBM	SP Power3
	P690
	Power5
	BG/L
SGI	Altix
HP Itanium-2	
QCDOC	
Intel / NVIDIA	w/ IB

Figure 1: On the left is a summary of domain application areas that were studied in this project. On the right is a short list of the machine architectures utilized through FY11.

1. Application Software Case Studies in FY04 for the Mathematical, Information and Computational Sciences Office of the U.S. Department of Energy, Roche KJ, J Drake, P Jones, DJ Dean, J Blondin, CP Ballance, MS Pindzola, C DeTar, J Osborn, R Brower, H Neff, and B Sugar. (2004), PNNL-24602, Pacific Northwest National Laboratory, Richland, WA.

2. Application Software Case Studies in FY05 for the Mathematical, Information and Computational Sciences Office of the U.S. Department of Energy, Roche KJ, EF Jaeger, LA Berry, DB Batchelor, E D'Azevedo, MD Carter , R Barrett, RW Harvey, RS Dumont, JR Myra, DA D'Ippolito, DN Smithe, CK Phillips, PT Bonoli, JC Wright, J Drake, P Worley, P Jones, S Plimpton, K Ko, N Folwell, L Ge, L Lee, Z Li, C Ng, G Schussman, R Uplenchwar, L Xiao, E Ng, W Gao, P Husbands, X Li, C Yang, G Golub, KL Ma, Z Bai, L Diachin, D Brown, P Knupp,

T Tautges, M Shephard, E Seol, D Keyes, O Ghattas, V Akcelik, JH Chen, E Hawkes, R Sankaran, J Sutherland, MR Fahey, D Skinner, HG Im, C Yoo, CJ Rutland, and AC Trouve. (2005), PNNL-24603, Pacific Northwest National Laboratory, Richland, WA.

3. Application Software Case Studies in FY06 for the Mathematical, Information and Computational Sciences Office of the U.S. Department of Energy, Roche KJ, T Schulthess, T Maier, M Norman, J Bordner, R Harkness, P Paschos, RJ Harrison, G Fann, R Hartman-Baker, W Shelton, S Sugiki, TP Straatsma, CS Oehmen, and J Nieplocha. (2006), PNNL-24604, Pacific Northwest National Laboratory, Richland, WA.

4. FY 2007 US OMB PMM DOE SC OASCR Software Metric SC GG 3.1/2.5.2: Improve Computational Science Capabilities, Roche KJ, A Mezzacappa, W Lee, JH Chen, J Blondin, S Bruenn, WR Hix, B Messer, M Adams, S Ethier, S Klasky, W Wang, E Hawkes, CK Law, D Lignell, T Lu, R Sankaran, C Yoo, JM Mellor-Crummey, S Shende, R Kendall, B de Supinski, and D Bailey. (2007), PNNL-24605, Pacific Northwest National Laboratory, Richland, WA.

5. FY 2008 US OMB PMM DOE SC OASCR Software Metric SC GG 3.1/2.5.2: Improve Computational Science Capabilities, Roche KJ, T Schulthess, M Eisenbach, T Maier, G Alvarez, PC Lichtner, RT Mills, G Hammond, J Candy, and MR Fahey. (2008), PNNL-24606, Pacific Northwest National Laboratory, Richland, WA.

6. FY 2009 US OMB PMM DOE SC OASCR Software Metric SC GG 3.1/2.5.2: Improve Computational Science Capabilities Roche KJ, D Kothe, R Kendall, S Ahern, J Hack, JC Oefelein, CS Chang, D Pugmire, TM Evans, H Childs, J Rosinski, KJ Evans, S Klasky, P Worley, E D'Azevedo, SH Ku, and R Sankaran. (2009), PNNL-24607, Pacific Northwest National Laboratory, Richland, WA.

7. FY 2010 US OMB PMM DOE SC OASCR Software Metric SC GG 3.1/2.5.2: Improve Computational Science Capabilities Roche KJ, D Kothe, R Hartman-Baker, R Kendall, A Bulgac, P Jones, M Maltrud, LW Wang, TM Evans, HA Nam, J White, P Worley, M Eisenbach, I Carpenter, and W Joubert. (2010), PNNL-24608, Pacific Northwest National Laboratory, Richland, WA.

8. FY 2011 US OMB PMM DOE SC OASCR Software Metric SC GG 3.1/2.5.2: Improve Computational Science Capabilities Roche KJ, P Crozier, S Plimpton, G Klimeck, M Luisier, S Steiger, W Mori, R Fonseca, SB Yabusaki, Y Fang, B Palmer, R Kendall, R Hartman-Baker, D Maxwell, B Messer, D Skinner, and V Paropkari. (2011), PNNL-24609, Pacific Northwest National Laboratory, Richland, WA.

There were many talks at meetings and conferences, but the relevant content is in the reports and has summarized results at the beginning of each report for each application. (Which means the great output of this project will remain a secret unless you bother to read the reports!)

APPENDIX

A GPRA-PMM Metrics

The U.S. Department of Energys Advanced Scientific Computing Research programs GPRA-PMM Software Metric for Computational Effectiveness is designed to comply with Public Authorizations PL 95-91, Department of Energy Organization Act, and PL 103-62, Government Performance and Results Act.
The U.S. Office of Management and Budget (OMB) oversees the preparation and administration of the Presidents budget; evaluates the effectiveness of agency programs, policies, and procedures; assesses competing funding demands across agencies; and sets the funding priorities for the federal government. The OMB has the power of audit and exercises this right annually for each federal agency. According to the Government Performance and Results Act of 1993 (GPRA), federal agencies are required to develop three planning and performance documents:

(a) Strategic Plan: a broad, 3 year outlook;

(b) Annual Performance Plan: a focused, 1 year outlook of annual goals and objectives that is reflected in the annual budget request (What results can the agency deliver as part of its public funding?); and

(c) Performance and Accountability Report: an annual report that details the previous fiscal year performance (What results did the agency pro- duce in return for its public funding?).

OMB uses its Performance Assessment Rating Tool (PART) to perform evaluations. PART has seven worksheets for seven types of agency functions. The function of Research and Development (RD) programs is included. RD programs are assessed on the following criteria:

- Does the RD program perform a clear role?
- Has the program set valid long term and annual goals?
- Is the program well managed?
- Is the program achieving the results set forth in its GPRA documents?

In Fiscal Year (FY) 2003, the Department of Energy Office of Science (DOE SC-1) worked directly with OMB to come to a consensus on an appro- priate set of performance measures consistent with PART requirements. The scientific performance expectations of these requirements reach the scope of work conducted at the DOE national laboratories. The Joule system emerged from this interaction. In FY 2008, the DOE transitioned from the Joule performance measure tracking system to OMBs Line of Business, Perfor- mance Measure Manager (PMM). The PMM is a performance-management database facilitated by the Treasury Department with the capability of up- loading performance metrics directly into OMBs PARTWeb system. The PMM organizes annual performance measures into various hierarchical struc- tures to show the relationship between individual performance targets and overall departmental performance. Departmental program and staff offices input performance measures and results directly into PMM on a periodic basis. This system is then used to produce the Performance Measure Details section of the DOEs Annual Performance Report that goes to Congress. In short, PMM enables the chief financial officer and senior DOE management to track annual performance on a quarterly basis. GPRA-PMM scores are reported as success, goal met (green light in PART), mixed results, goal partially met (yellow light in PART), and unsatisfactory, goal not met (red light in PART). GPRA-PMM links the DOE strategic plan5 to the underlying base program targets.

B GPRA-PMM Goals for the DOE ASCR Program

Below is an example of ASCR's annual goals. This project focused on goal 2.

(a) **SC GG 3.1/2.5.1** Focus usage of the primary supercomputer at the National Energy Research Scientific Computing Center (NERSC) on capability computing, defined as the percentage of the computing time used by computations that require at least 1/8 of the total resource. FY11 performance metric: capability usage is at least 40%.

(b) **SC GG 3.1/2.5.2** Improve computational science capabilities, defined as the average annual percentage increase in the computational effectiveness (either by simulating the same problem in less time or simulating a larger problem in the same time) of a subset of application codes. FY11 performance metric: efficiency measure is 100%.

Ensuring compliance with these metrics, which are tracked on a quarterly basis, was an important milestone each fiscal year for the DOE ASCR Program Office as well as for the success of the overall DOE SC-1 open science computing effort. The reports detail the results of the effectiveness of the computational science capability (SC GG 3.1/2.5.2) each FY listed.

C Quarterly Tasks Related to SC GG 3.1/2.5.2

The GPRA-PMM effort to improve computational science capabilities is a year long effort requiring quarterly updates. The quarterly sequence of tasks for exercising this software metric is as follows.

Quarter One (Q1) Tasks (deadline: December 31). Identify a sub- set of candidate applications (scientific software tools) to be investigated on DOE SC supercomputers. Management (at DOE SC and national laboratories) decides upon a short list of applications and computing platforms to be exercised. The Advanced Scientific Computing Advisory Committee (AS- CAC) approves or rejects the list. The Q1 milestone is satisfied when a short list of target applications and machines (supercomputers) is approved.

Quarter Two (Q2) Tasks (deadline: March 31). Problems that each chosen application will simulate on the target machines are determined. The science capability (simulation result) and computational performance of the implementation are benchmarked and baselined (recorded) on the target ma- chines for the defined problems and problem instances. The Q2 milestone is satisfied when benchmark data is collected and explained. If an application is striving to achieve a new science result in addition to demonstrating improved performance, then providing a detailed discussion of the existing capability, why it is insufficient and how the challenges / deficiencies are being addressed satisfy the Q2 milestone.

Quarter Three (Q3) Tasks (deadline: June 30). The application soft- ware (its models, algorithms, and implementation) is enhanced for efficiency, scalability, science capability, etc. The Q3 milestone is satisfied when the status of each application is reported at the Q3 deadline. Corrections to Q2 problem statements are normally submitted at this time.

Quarter Four (Q4) Tasks (deadline: September 30). Enhancements to the application software continue as in Q3. The enhancements are stated and demonstrated on the machines used to generate the Q2 baseline information. A comparative analysis of the Q2 and Q4 data is summarized and reported. The Q4 milestone is satisfied if the enhancements made to the application software are in accordance with the efficiency measure as defined in Q2 (run-time efficiency, scalability, or new result).